Draft version January 31, 2013
Preprint typeset using LaTeX style emulateapj v. 5/2/11

OPENING THE 21CM EOR WINDOW: MEASUREMENTS OF FOREGROUND ISOLATION WITH PAPER

Jonathan C. Pober[1], Aaron R. Parsons[1], James E. Aguirre[2], Zaki Ali[1], Richard F. Bradley[3,4,5], Chris L. Carilli[6], Dave DeBoer[7], Matthew Dexter[7], Nicole E. Gugliucci[5], Daniel C. Jacobs[8], Dave MacMahon[7], Jason Manley[9], David F. Moore[2], Irina I. Stefan[10], William P. Walbrugh[9]

Draft version January 31, 2013

ABSTRACT

We present new observations with the Precision Array for Probing the Epoch of Reionization (PAPER) with the aim of measuring the properties of foreground emission for 21cm Epoch of Reionization experiments at 150 MHz. We focus on the footprint of the foregrounds in cosmological Fourier space to understand which modes of the 21cm power spectrum will most likely be compromised by foreground emission. These observations confirm predictions that foregrounds can be isolated to a "wedge"-like region of 2D $(k_\perp, k_\parallel)$-space, creating a window for cosmological studies at higher $k_\parallel$ values. We also find that the emission extends past the nominal edge of this wedge due to inherent spectral structure in the foregrounds themselves, with this feature most prominent on the shortest baselines. Finally, we filter the data to retain only this "unsmooth" emission and image it. The resultant image shows an excess of power on large angular scales, but no emission can be clearly localized to any one region of the sky. This image is highly suggestive that the most problematic foregrounds for 21cm EoR studies will not be easily identifiable bright sources, but rather an aggregate of fainter emission.

Subject headings: cosmology: observations — dark ages, reionization, first stars — techniques: interferometric

1. INTRODUCTION

The highly redshifted 21cm line of neutral hydrogen is widely regarded as one of the most promising probes of the high redshift universe, with potential to map out volumes extending from redshift ~ 1 through the Epoch of Reionization (EoR) to the dark ages at redshift 20 and beyond (for reviews of the field, see Furlanetto et al. 2006, Morales & Wyithe 2010, and Pritchard & Loeb 2012). Numerous facilities and experiments targeting the signal from the EoR are already online or under construction, including the LOw Frequency ARray (LOFAR; Yatawatta et al. 2013)[11], the Murchison Widefield Array (MWA; Tingay et al. 2012)[12], and the Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER; Parsons et al. 2010)[13]. All 21cm cosmology experiments will need to separate bright galactic and extragalactic foregrounds from the neutral hydrogen signal, which can be fainter by as much as 5 orders of magnitude or more (see, e.g., Furlanetto et al. 2006 and Santos et al. 2005).

Considerable effort has been devoted to developing schemes to remove or isolate foregrounds from 21cm data (e.g. Morales et al. 2006, Bowman et al. 2009, Liu et al. 2009, Liu & Tegmark 2011, Parsons et al. 2012b, Dillon et al. 2012). Almost all of these approaches rely on the spectral smoothness of foreground emission relative to the 21cm signal, which will contain significant structure versus frequency. The purpose of this letter is to use the delay transform technique presented in (Parsons et al. 2012b; hereafter, P12b) on observations from PAPER to test the behavior of actual foreground emission. We wish to understand the footprint of foregrounds in k-space to determine which modes of the 21cm power spectrum will be most accessible to observation. The structure of this letter is as follows: in §2, we describe the data used in these observations. In §3, we review the delay spectrum technique presented in P12b, and then describe the steps used in applying this approach to actual observations. We present our results in §4 and conclude in §5.

2. THE DATA

We use 4 hours of data collected in 10 second integrations between JD 2455747.48 and 2455747.64 ($4-5$ July 2011), using a 64 element PAPER array located on the SKA site in the Karoo region of South Africa. This data set comes from the same observing campaign described by Stefan et al. (2012), although this specific 4-hour window falls outside of the observations analyzed therein. The dipole antennas are arranged in a "minimum redundancy" configuration optimized for imaging analysis (Parsons et al. 2012a), which is shown in Figure 1. This configuration has a maximum baseline length of ~ 300m, corresponding to a maximum image plane resolution of 0.4° at 150 MHz. The PAPER correlator has a 100 MHz instantaneous bandwidth from $100-200$ MHz divided into 2048 frequency channels. We correlate only one linear polarization on each dipole, and discard all data from antennas 40 and 55 which were cross polar-

[1] Astronomy Dept., U. California, Berkeley, CA
[2] Dept. of Physics and Astronomy, U. Pennsylvania, Philadelphia, PA
[3] Dept. of Electrical and Computer Engineering, U. Virginia, Charlottesville, VA
[4] National Radio Astronomy Obs., Charlottesville, VA
[5] Dept. of Astronomy, U. Virginia, Charlottesville, VA
[6] National Radio Astronomy Obs., Socorro, NM
[7] Radio Astronomy Lab., U. California, Berkeley, CA
[8] School of Earth and Space Exploration, Arizona State U., Tempe, AZ
[9] Square Kilometer Array, South Africa Project, Cape Town, South Africa
[10] Cavendish Lab., Cambridge, UK
[11] http://www.lofar.org/
[12] http://www.mwatelescope.org/
[13] http://eor.berkeley.edu/

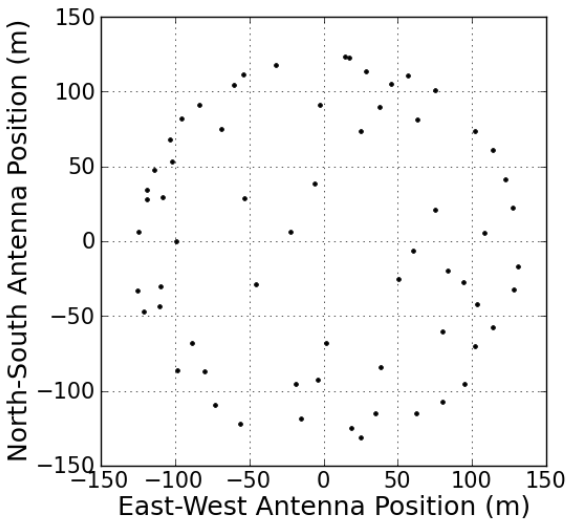

Figure 1. The configuration of the 64 PAPER dipoles used in this analysis. The zero-point is the center of the array. The y-axis is North/South, the x-axis is East/West, and distances are in meters.

ized during the time of the observation.

An image of the field transiting during this time period is shown in Figure 2. This image spans a 25 MHz band spanning $140 - 165$ MHz; 100 sub-bands of 0.25 MHz were individually imaged and summed to make the map shown. No CLEANing or spectral slope correction was performed, and the visibilities are weighted naturally. The observation is centered on a low-foreground "cold patch," a potential field for an EoR science observation. The Galactic plane, which is just setting at the end of the 4-hour observation, creates prominent sidelobes over the entire map.

Complex antenna based gains are derived from fringe-fitting to Centaurus A, Pictor A, and Fornax A; an overall gain scale is derived from the Helmboldt et al. (2008) source J2214-170. The details of this calibration will be presented in a forthcoming paper by Parsons et al. We also perform minimal preparatory processing on the data for the power spectrum analysis. These steps include a small gain linearization correction to mitigate data quantization effects in the correlator (described in Parsons et al. 2010) and a correction for temperature dependent gain drifts (described in Pober et al. 2012). We then perform a radio-frequency interference (RFI) excision, manually flagging frequency channels of known transmitters, and flagging any points 6σ above the mean after differencing adjacent channels along the frequency axis.

3. ANALYSIS TECHNIQUES

At the core of our analysis is the delay spectrum technique presented in P12b. Since cosmological redshifting maps the observed 21cm line frequency into a distance measurement, the Fourier transform of the frequency axis are the $k_\parallel$ line-of-sight modes of the 21cm power spectrum. However, the frequency dependent sampling pattern of an interferometric baseline also turns the frequency Fourier transform of a single baseline's visibility spectrum into the "delay transform" first presented in Parsons & Backer (2009). The delay transform maps sources of celestial emission to Dirac delta functions in delay space corresponding to the geometric delay in arrival time between the two elements of a baseline.

The power of the delay transform in the cosmological context is the framework it provides for understanding the area of k-space affected by foreground emission. Since geometric delays are limited to values below the physical light travel time distance between the two antennas (the "horizon limit"), all emission from the sky maps to a small region in the center of delay space. However, any other spectral structure in the visibilities acts as a convolving kernel for the celestial emission. Foreground emission is spectrally smooth, translating into a narrow convolving kernel; 21cm emission has large amounts of spectral structure, and therefore its kernel scatters power from within the horizon limit to larger delays. P12b also showed that delay has a near one-to-one mapping to $k_\parallel$, meaning both that foreground emission is confined to large scale modes along the line of sight and that those delay modes free from contaminating foreground emission are effective probes of the 21cm power spectrum.

3.1. *Delay Space CLEAN*

Some of the practical aspects of implementing the delay-spectrum approach in actual data were described in §3 of P12b. Of particular importance is the implementation of the frequency Fourier transform using a Blackman-Harris window function and the 1D-CLEAN algorithm first presented in Parsons & Backer (2009) to reduce the effects of RFI flags and band-edge effects. Even if foregrounds are spectrally smooth and easily localized in delay/$k_\parallel$-space, such sharp edges in frequency space will scatter power from foregrounds into otherwise uncontaminated regions of k-space.

The 1D CLEAN algorithm treats RFI flags as a sampling function in frequency space, and "fills in" these gaps by iteratively fitting the brightest Fourier components in the delay domain. We force our model to be smooth-spectrum in frequency by only allowing delay-space components which fall inside a baseline-dependent CLEAN area (a "CLEAN box"), 50 ns beyond the physical maximum horizon delay on that baseline. This extra 50 ns allows CLEAN to model foreground emission pushed beyond the horizon limit. The exact value of 50 ns is chosen to minimize the amount of foreground emission remaining in the residuals, so that the sidelobes of any remaining flux scattering off RFI gaps are below the noise. (While the foreground "model" is free of RFI gaps, the residuals will not be.) The analysis below does reveal emission extending more than 50 ns beyond the horizon limit, and P12b predicts that at the level of the EoR signal, foreground emission can extend as much as 300 ns beyond the horizon. At the sensitivity level of this work, however, we find 50 ns to suppress the sidelobes of foreground emission in our "EoR" window to the noise level. The exact value of the cut-off chosen will need to be revisited in future analyses with more sensitivity.

The end result of CLEAN is a smooth-spectrum model of our data which is free of RFI flagging gaps. We also retain a data-set consisting of the residuals between this model and the raw data to ensure that no information is lost.

3.2. *Power Spectra*

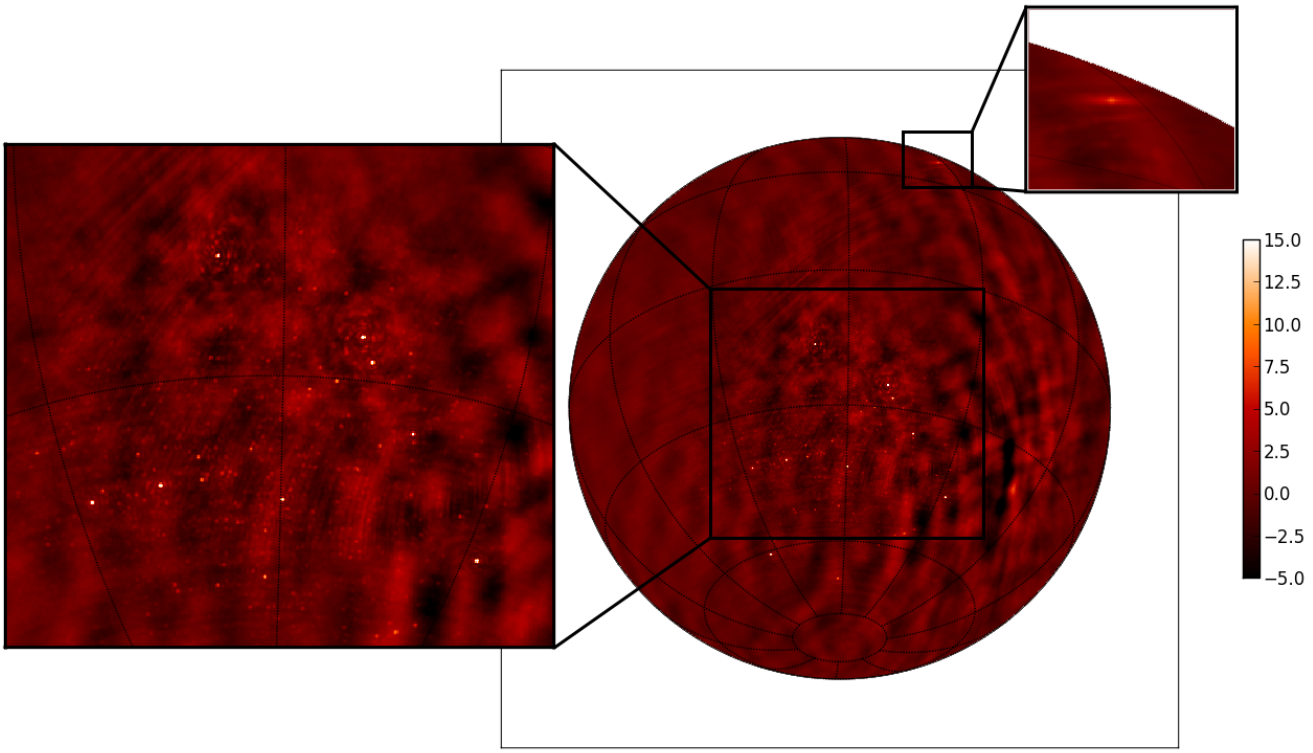

Figure 2. A dirty image of the data used in this analysis, centered on RA 21h52m and declination -30°43', the transiting zenith halfway through the observation. Prominent sidelobes from the Galaxy are seen in the right-hand side of the map. Close-ups show Cygnus A at 19h59m and +40°44' and the point source population of a potential EoR cold-patch. The color-scale is linear in Jy, with only the brightest point sources saturating the range.

Once the data have been CLEANed, is it quite straightforward to form power spectra on a visibility-by-visibility basis. Our power spectrum estimates follow from equation (12) of Parsons et al. (2012a):

$$\widehat{P}(k) \approx \tilde{V}_{21}^2 \left(\frac{\lambda^2}{2k_{\rm B}}\right)^2 \frac{X^2 Y}{\Omega B}, \qquad (1)$$

where λ is the observing frequency, $k_{\rm B}$ is Boltzmann's constant, Ω is the solid angle of the primary beam,[14] B is the observing bandwidth, X and Y are cosmological scalars which convert observed angles and frequencies into $h{\rm Mpc}^{-1}$, and $\tilde{V}$ is a delay transformed visibility. We avoid introducting a noise-bias by forming our estimator $\tilde{V}_{21}^2$ from adjacent time samples on a baseline. The 10-second interval between integrations is short enough that both measurements can be considered redundant samples of the same k-modes.

The isotropy of the universe allows us to then combine all power spectrum estimates $\tilde{V}_{21}^2$ in annuli of equal $k_\perp$ to form a 2D power spectrum in the $(k_\perp, k_\parallel)$-plane. We note that the method used here does not take advantage of any coherent integration within a uv-pixel. Since foreground emission dominates the observed signal, the loss of sensitivity is not important, and we ignore this effect for computational efficiency.

[14] Derivations of equation 1 in Morales (2005), McQuinn et al. (2006), and Parsons et al. (2012a), relate ΩB to an effective cosmological volume, using a top-hat primary beam or effective area as a pedagogical tool to simplify the result. More generally, however, the effective Ω in equation 1 is actually $\int A^2(\theta,\phi)\, {\rm d}\theta\, {\rm d}\phi$, where $A(\theta,\phi)$ is the power response of the primary beam. A derivation of this effect will be presented in a subsequent full length publication by Parsons et al. For PAPER, this beam is a factor of ~ 2 smaller than $\int A(\theta,\phi)\, {\rm d}\theta\, {\rm d}\phi$.

4. RESULTS

To make power spectra of the 4-hour dataset described above, we first run our CLEAN algorithm over the full 100 MHz band. Using the whole band gives the best resolution in delay space, and since foregrounds are nearly coherent over the whole band, the additional information gives CLEAN the most signal-to-noise to build its model. For the cosmological delay transform to make power spectra, we use only a 25 MHz band from 140 to 165 MHz. This smaller band still exceeds the ~ 8 MHz band over which the $z \sim 10$ universe can be treated as coeval (Furlanetto et al. 2006). Since the main purpose of this work is to understand the k-space behavior of foregrounds, we ignore this effect, as the additional bandwidth gives us better k-space resolution. For an EoR analysis, the isolation of foregrounds will be slightly worse than shown here, owing to the necessarily poorer resolution.

Forming individual power spectra from each baseline of the array and binning in annuli of constant $|k_\perp|$ yields the 2-dimensional $P(k_\perp, k_\parallel)$ shown in Figure 3. The binning in $|k_\perp|$ is very minimal; most bins contain only one baseline and gaps are still visible (especially at the small $k_\perp$-values) where there are no baselines of that length.

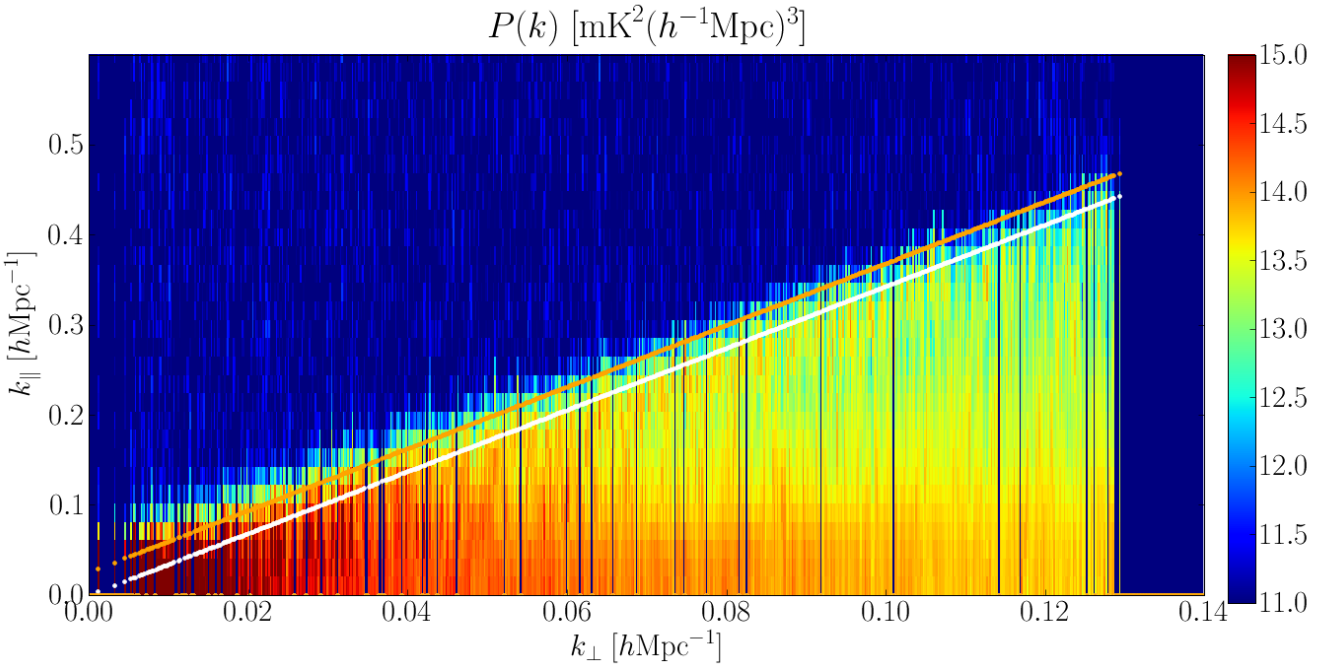

Figure 3. A two-dimensional power spectrum of the 4 hours of data analyzed. The wedge-like nature of the foreground emission is extremely prominent. The white line marks the horizon limit and the orange line is 50 ns beyond. Binning is minimal; each $k_\perp$ values corresponds to at most a few baselines, with gaps caused by a lack of a baseline with that magnitude. The "edge brightening" of the foreground emission on the longest baselines as one moves near the horizon limit can be attributed to the Galactic center, and moves as expected when the data are viewed as a function of time.

We do no binning in $k_\parallel$; this resolution is set by the 25 MHz bandwidth used in the analysis.[15] The most prominent feature is the "wedge"-like shape of the foreground emissison as predicted in a number of studies, including Datta et al. (2010), Vedantham et al. (2012), Morales et al. (2012), Trott et al. (2012), and P12b. As argued in P12b, this "wedge" footprint in k-space is not a result of imperfect calibration in foreground removal (since we attempt no foreground removal in this work), but a property of the emission itself as measured by an interferometer.

Also of note are the two diagonal lines plotted on top of the power spectrum. The white line corresponds to the horizon-limit in $k_\parallel$ for a baseline of corresponding length $k_\perp$; the orange line is 50 ns beyond the horizon, corresponding to the CLEAN region used in our deconvolution. As predicted, emission extends beyond the horizon limit due the intrinsic spectral structure of the foreground emission.

We draw attention to the fact that the supra-horizon emission does not have a constant width in $k_\parallel$ as a function of $k_\perp$. Rather, more emission extends beyond the horizon on the smallest $k_\perp$-values (i.e. the shortest baselines). The likeliest explanation for this result is a combination of two different effects. First, the shortest baselines will "see" more total emission from the sky, since they resolve out less of the diffuse Galactic synchrotron. Therefore, the sidelobes of this emission will be brighter as well, so that we can see their extent further in $k_\parallel$ before they fall below the noise level. The second effect is somewhat more subtle. A given amount of spectral "unsmoothness" corresponds to a kernel with a fixed width in delay (and therefore, $k_\parallel$) space. However, the horizon limit moves to higher $k_\parallel$ values for longer baselines. Therefore, it is easier for a fixed width kernel of unsmooth emission to corrupt supra-horizon delays on shorter baselines than on longer.

To highlight the steepness of the foreground roll-off, we plot a 1-dimensional $k_\parallel$ power spectrum for bins of several baseline lengths in Figure 4. We see that the foreground emission can fall by as much as three to four orders of magnitude in a factor of 2 change in $k_\parallel$. It is difficult to explicitly compare this result to the predictions of P12b, due to the different resolutions and binning used. The placement of bin edges can significantly complicate comparison when the fall-off is so steep, as a slight shift in the bin can result in a large change in the average value within. For similar reasons, it is difficult to say exactly where the foreground fall-off falls below the noise. Given these caveats, there is nothing in this data to contradict the predictions of P12b. It is clear, however, that sensitivities will have to increase several more orders of magnitude before anything can be said about the behavior of foreground emission at the tens of mK^2 level where the expected EoR signal lies.

To further investigate the nature of the supra-horizon emission, we filter the data in delay-space, selecting only the delay bin just beyond the 50 ns cutoff. That is to say, we image the emission from the bin lying just above the orange line in Figure 3.[16] The resultant map is shown in Figure 5. None of the bright point sources in Figure 2 remain visible, nor does any emission associated with the Galactic plane. Closer inspection reveals low spatial-

[15] Because of the Blackman-Harris window function used in CLEAN, only every other sample plotted in $k_\parallel$ is statistically independent.

[16] While we might ideally want an image of *all* the emission above the horizon-limit, a full data-cube versus frequency would be required to observe all the different $k_\parallel$ modes. Given the short observation time, this cube is too noisy; instead, we image only one mode per-baseline, which allows us to take advantage of the full band's sensitivity.

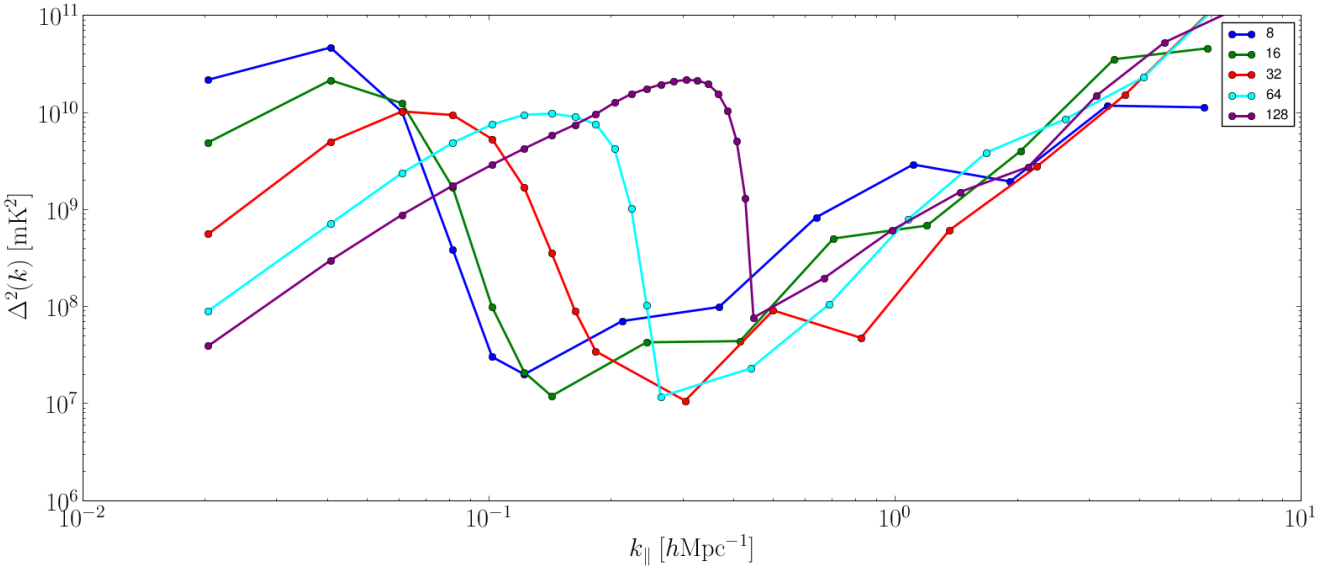

Figure 4. A 1-dimensional power spectrum versus $k_\parallel$ for bins of several baseline lengths. Labels correspond to the maximum length baseline in the bin, e.g., the 16-wavelength bin contains baselines from 8 to 16 wavelengths. To preserve the steep roll-off of the foreground emission, the data are plotted with their natural resolution at lower $k_\parallel$ values, and logarithmically binned at higher values.

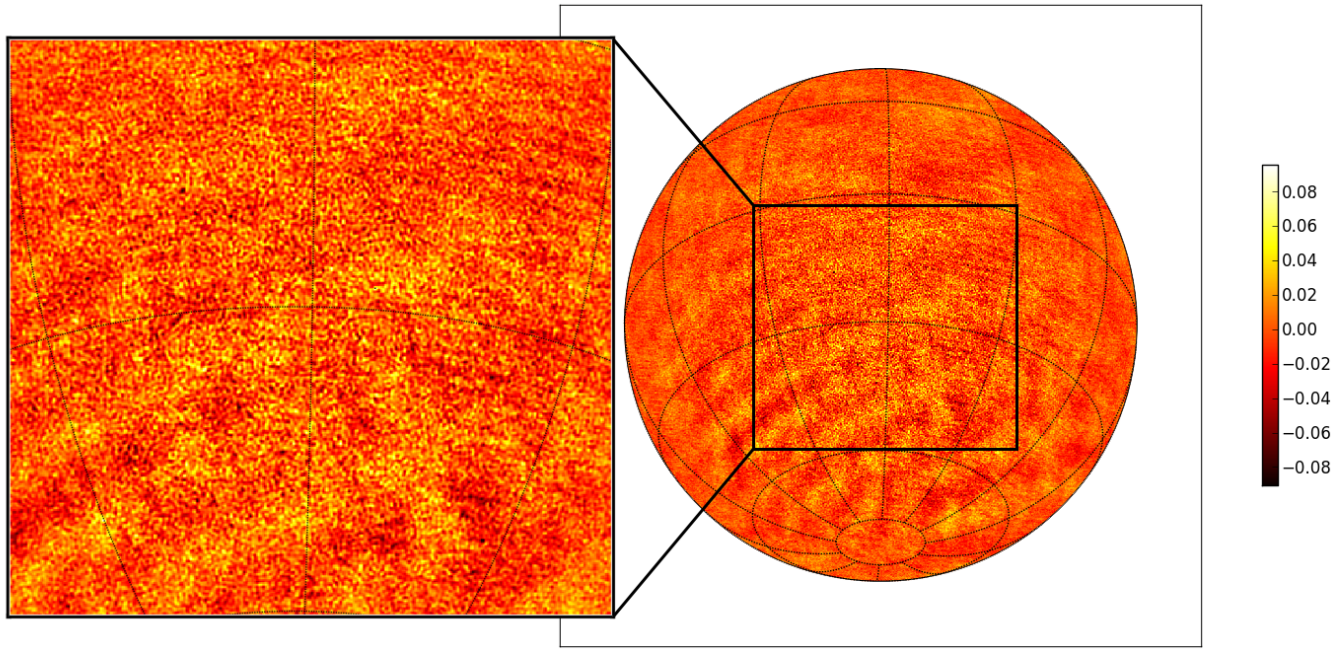

Figure 5. A dirty image of the data, after selecting only th delay bin just past 50 ns beyond the horizon limit (i.e. one bin beyond the orange line in Figure 3). The color-scale is linear in Jy. The dominant feature in the map is an excess of emission above the noise on large spatial scales. No emission can be identified with the bright sources in Figure 2, nor are the large scale features obviously correlated with Galactic structure.

frequency ripples across the map.[17]

We can confirm the existence of this excess power by comparing the rms of the image to an estimate of the expected noise level in the map. Using parameters for a 4 hour observation at 150 MHz with a 25 MHz bandwidth and 62 antennas gives:

$$\Delta\sigma = \frac{2k_B\Omega}{\lambda^2}\frac{T_{\rm sys}}{\sqrt{N(N-1)Bt}} \approx 14 \text{ mJy}, \qquad (2)$$

where k_B is Boltzmann's constant, Ω is the solid angle of the primary beam, λ is the observing wavelength, $T_{\rm sys}$ is the system temperature, N is the number of antenna in the array, B is the bandwidth, and t is the observing time

[17] The use of natural weighting as in Figure 2 will lead to a scale dependent noise across the image; to correct for this effect, Figure 5 was made using uniform weighting of the visibilities, and the low spatial-frequency power is still visible.

(Thompson et al. 2007); fiducial values for PAPER are $\Omega = 0.75$ sr and $T_{\rm sys} = 1000$ K. The exact value of $T_{\rm sys}$ is likely too high by as much as a factor of ~ 2, but in this case, using a high value for $T_{\rm sys}$ makes for a conservative estimate. A more thorough discussion of $T_{\rm sys}$ for PAPER will be presented in a forthcoming publication by Parsons et al. The measured rms from $\sim 15,000$ independent measurements near the center of Figure 5 (where primary beam effects can be ignored) is 21 mJy, which supports our hypothesis that the structure in Figure 5 is real.

Given the $k_\perp$ dependence of the emission above the orange line in Figure 3, it is not surprising that excess power is seen on large angular scales. We have already argued above why we might expect such a result in the power spectrum domain. What is surprising is that none of this emission can be easily associated with the brightest structures visible in the image domain. It appears that the bulk of the emission contaminating the EoR window comes not from easily localizable sources or the Galactic plane itself, but from an aggregate of fainter emission. However, we caution against over-interpretation of Figure 5; some or all of this supra-horizon emission could be created by frequency structure in the PAPER instrument itself. The frequency structure in both simulated and calibrated models of the PAPER system (Parsons et al. 2010; Pober et al. 2012) would be a negligible factor in pushing smooth-spectrum emission beyond the horizon (P12b), but we cannot rule out the possibility that the instrument has additional yet-unmeasured frequency structure.

5. CONCLUSIONS

We have presented new observations from PAPER measuring the properties of foreground emission in cosmological Fourier space. These observations have confirmed general predictions presented in, e.g., Datta et al. (2010), Morales et al. (2012) and P12b: that foreground emission occupies a "wedge" in the 2D $(k_\perp, k_\parallel)$ plane, leaving a window at higher $k_\parallel$ values for 21cm EoR studies. We have also confirmed that shorter baselines yield a larger window onto the cosmological signal. However, this the window does not grow perfectly linearly with decreasing baseline length. A combination of the increased brightening of diffuse Galactic synchrotron on the shortest baselines and the size of a given delay-space kernel relative to sky-delays both lead to excess encroachment past the nominal "horizon-limit". Therefore, while shorter baselines do make the best probes of the EoR signal, there will be diminishing returns at the shortest baselines.

We have also presented an image of the supra-horizon "unsmooth" emission extending past the nominal edge of the wedge. This image confirms the existence of excess power on the shortest baselines in the array, but is unable to localize any of the emission to known sources on the sky. This result is highly suggestive that a targeted removal of the brightest sources of emission may have only a minor effect on increasing the size of the EoR window.

We thank the SKA project office in South Africa for their efforts in ensuring the smooth running of PAPER. The PAPER project is supported through the NSF-AST program (award 0804508), and by significant efforts by staff at NRAOs Green Bank and Charlottesville sites.

REFERENCES

Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2009, ApJ, 695, 183
Datta, A., Bowman, J. D., & Carilli, C. L. 2010, ApJ, 724, 526
Dillon, J. S., Liu, A., & Tegmark, M. 2012, arXiv:1211.2232
Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, Phys. Rep., 433, 181
Helmboldt, J. F., Kassim, N. E., Cohen, A. S., Lane, W. M., & Lazio, T. J. 2008, ApJS, 174, 313
Liu, A., & Tegmark, M. 2011, Phys. Rev. D, 83, 103006
Liu, A., Tegmark, M., Bowman, J., Hewitt, J., & Zaldarriaga, M. 2009, MNRAS, 398, 401
McQuinn, M., Zahn, O., Zaldarriaga, M., Hernquist, L., & Furlanetto, S. R. 2006, ApJ, 653, 815
Morales, M. F. 2005, ApJ, 619, 678
Morales, M. F., Bowman, J. D., & Hewitt, J. N. 2006, ApJ, 648, 767
Morales, M. F., Hazelton, B., Sullivan, I., & Beardsley, A. 2012, ApJ, 752, 137
Morales, M. F., & Wyithe, J. S. B. 2010, ARA&A, 48, 127
Parsons, A., Pober, J., McQuinn, M., Jacobs, D., & Aguirre, J. 2012a, ApJ, 753, 81
Parsons, A. R., & Backer, D. C. 2009, AJ, 138, 219
Parsons, A. R., Pober, J. C., Aguirre, J. E., Carilli, C. L., Jacobs, D. C., & Moore, D. F. 2012b, ApJ, 756, 165
Parsons, A. R., et al. 2010, AJ, 139, 1468
Pober, J. C., et al. 2012, AJ, 143, 53
Pritchard, J. R., & Loeb, A. 2012, Reports on Progress in Physics, 75, 086901
Santos, M. G., Cooray, A., & Knox, L. 2005, ApJ, 625, 575
Stefan, I. I., et al. 2012, arXiv:1212:1624
Thompson, A. R., Moran, J. M., & Swenson, G. W. 2007, Interferometry and Synthesis in Radio Astronomy, John Wiley & Sons, 2007.
Tingay, S. J., et al. 2012, arXiv:1206:6945
Trott, C. M., Wayth, R. B., & Tingay, S. J. 2012, ApJ, 757, 101
Vedantham, H., Udaya Shankar, N., & Subrahmanyan, R. 2012, ApJ, 745, 176
Yatawatta, S., et al. 2013, arXiv:1301:1630